基于 BiGRU-Attention 神经网络的文本情感分类模型 *

王 伟[a,b]，孙玉霞[b]，齐庆杰[c]，孟祥福[b]

(辽宁工程技术大学 a. 科学技术研究院; b. 电子与信息工程学院; c. 矿业学院，辽宁 葫芦岛 125105)

摘 要：针对双向长短时记忆神经(BiLSTM)模型训练时间长、不能充分学习文本上下文信息的问题，提出一种基于 BiGRU-Attention 的文本情感分类模型。首先，利用双向门控循环(BiGRU)神经网络层对文本深层次的信息进行特征提取；其次，利用注意力机制(attention)层对提取的文本深层次信息分配相应的权重；最后，将不同权重的文本特征信息放入 softmax 函数层进行文本情感极性分类。实验结果表明，所提的神经网络模型在 IMDB 数据集上的准确率是 90.54%，损失率是 0.2430，时间代价是 1100 s，验证了 BiGRU-Attention 模型的有效性。

关键词：文本情感分类；注意力机制；双向门控循环神经网络

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2018.07.0413

Text sentiment classification model based on BiGRU-Attention neural network

Wang Wei[a, b], Sun Yuxia[b], Qi Qingjie[c], Meng Xiangfu[b]

(*a. College Institute of Science & Technology, b. College of Electronic & Information Engineering, c. College of Mining Liaoning Technical University, Huludao Liaoning* 125105, *China*)

Abstract: The BiLSTM neural network model has long training time and can not fully learn text context informations. In order to solve the problems, this work proposes a text emotion classification model based on BiGRU-Attention neural network. Firstly, the bidirectional Gated Recurrent Unity (BiGRU) neural network layer was used to extract the features of the deep text information; secondly, the attention mechanism (Attention) layer was used to allocate the corresponding weight of the extracted text deep information. Finally, the text feature messages of different weights are put into the softmax function layer to carry out the text sentiment classifications. The experimental results show that the accuracy of the proposed neural network model is 90.54% on the IMDB data set, the loss rate is 0.2430 and the time cost is 1100s and the validity of the BiGRU-Attention model is verified.

Key words: text sentiment classifications; attention mechanism; bidirectional Gated Recurrent Unity (BiGRU)

0　引言

文本情感分类是 NLP 领域的一个重要任务，通常是对带有主观情感色彩的文本进行分类。文本情感分类能对文本层次特征进行提取，挖掘用户的情感倾向，被广泛应用在政策意见挖掘、民意调查、产品分析、电影推荐、搜索排名等领域。近年来，Web 2.0 的兴起掀起了用户主动创造信息的时代，用户可以随时通过移动终端发表自己意见和评论，因此用户评论数据呈指数型增长。研究人员最初使用的朴素贝叶斯等浅层次机器学习算法已不能满足与日俱增的数据处理的需要，因此神经网络应运而生并且成为近年来的研究热点。

CNN 最初在图像领域得到广泛应用，并迅速扩展到其他领域。LeCun 等人[1]将 CNN 应用到文本情感分类，提高了文本分类的准确率。虽然采用 CNN 相比传统研究人员所采用的基于规则和机器学习[2]算法准确率提高了，但仍不能充分学习文本上下文信息。随着研究的不断深入，Mikolov 等人[3]提出将 RNN 应用到文本分类任务，由于 RNN 当前节点的输出值由当前的输入和上一个节点的输出两部分共同决定，使得 RNN 能够充分学习文本前后上下之间的信息，因此 RNN 比 CNN 适合于文本分类；但 RNN 长期依赖学习的特征会容易出现梯度弥散的问题。为了解决梯度弥散问题，LSTM 和 GRU 等众多变体被提出并在情感分类领域广泛应用，但 LSTM 由于结构的复杂性，计算异常复杂，存储了多余的中间变量，因此需要大量的训练时间和内存空间且 LSTM、GRU 只能利用历史信息来对当前信息作出判断，无法利用未来信息，有时可能无法对当前信息进行准确的判断和充分地文本信息提取；目前广泛在文本领域应

收稿日期：2018-07-23；修回日期：2018-09-11　基金项目：国家自然科学基金资助项目（61772249）；2016 年辽宁省教育厅城市研究院一般项目（LJCL008）

作者简介：王伟（1981-），男，安徽砀山人，副教授，博士研究生，主要研究方向为机器学习、大数据分析与查询（39846661@qq.com）；孙玉霞（1992-），女，硕士研究生，主要研究方向为神经网络、自然语言处理；齐庆杰（1964-），男，辽宁朝阳人，教授，博导，主要研究方向为矿山安全大数据分析、矿井灾害防治；孟祥福（1981-），男，辽宁朝阳人，副教授，博导，主要研究方向为大数据分析与查询、时空数据挖掘与机器学习算法.

用的 BiLSTM 和 attention 机制结合的模型，仍无法摆脱复杂计算造成的训练时间过长的问题。

针对上述问题，本文提出了一种 BiGRU-Attention 模型，将文本的输入向量放入 BiGRU 层相比放入 BiLSTM 层，能进一步减少网络训练的时间；把 BiGRU 层的输出向量放入 Attention 层能进一步突出文本的关键信息，提高文本信息的提取质量。BiGRU-Attention 模型在 IMDB 数据集上进行对比实验，将准确率、损失率和时间代价作为评价指标，证明了 BiGRU-Attention 模型在文本分类方面的有效性。

1　相关研究

利用深度学习的方法研究文本情感分类近年如雨后春笋般的涌现。梁军等人[4]采用递归神经网络的情感转移模型加强文本相关性的捕获；白静等人[5]提出使用 BiLSTM-CNN-Attention 的混合神经网络结构进行两种特征融合的分类；梁斌等人[6]采用 attention 机制和 CNN 结合的混合模型，解决平行化问题，降低模型训练时间；赵勤鲁等人[7]提出 LSTM-Attention 模型充分提取了语义结构信息；杨东等人[8]提出 CNN、GRU 和 Attention 机制的混合神经网络模型对文本进行特征提取；司念文等人[9]提出 attention 机制和 LSTM 混合模型更加有效地对中文词性进行标注；王业沛等人[10]采用 LSTM 神经网络进行判断结果的倾向性任务分析，有效地提高了裁判文书中关键信息提取的准确性；朱星嘉等人[11]提出了改进的 attention-based LSTM 的特征选择模型，解决了数据的维度灾难问题，有效地突出了文本的重点特征信息；李阳辉等人[12]将降噪自编码的深度学习方法应用在情感分析任务，提高了模型对原始数据的鲁棒性和信息特征的表达能力；刘洋[13]采用 GRU 神经网络对时间序列进行预测，提高了预测方案的准确性；李骁等人[14]采用 BiGRU 网络对互联网信息输入序列快速准确地实现特定信息的提取；黄兆玮等人[15]采用 GRU 和 attention 机制结合的远程监督关系抽取的方法提高了准确率；张玉环等人[16]利用 LSTM 及 GRU 搭建文本情感分类模型使模型在较短的时间内达到较高的准确率；田竹[17]将 RSGRU 的混合神经网络应用在语句级情感分析任务，节约了人力，更加容易维护； Zhou 等人[18]采用最小门控单元(MGU)对序列化的数据进行处理，加快了模型的训练速度；黄磊等人[19]采用 LSTM 和 GRU 的混合神经网络模型对文本重点信息进行提取，召回率明显提高。

近年，attention 机制在自然语言处理领域广泛应用，attention 机制最早应用在计算机视觉图像领域。2014 年 google mind 团队[20]使用循环神经网络和 attention 机制的混合模型做图像分类，它的主要思想是当观察某一幅图片时，人们经常会把注意力集中到某一小块区域，并能根据之前的观察来预测未来的注意力应该集中图像的位置，根据当前的输入和前一状态去确定当前状态应该处理的注意力部分，处理的像素更少，问题任务更加简单化；随后 Bahdanau 等人[21]将 attention 机制应用在机器翻译任务，通过 attention 机制把源语言端每个词学到的表达和预测要翻译的词联系起来，这也是 attention 机制在自然语言处理领域的首次应用；Luong 等人[22]介绍了全局和局部两种注意力机制；Yin 等人[23]介绍了卷积和 attention 机制结合的三种方式，可以在输入 CNN 之前加入 attention 机制，可以在进行 CNN 特征提取后，池化之前加入 attention 机制；也可以把以上两种方式结合起来添加 attention 机制，这也是 attention 机制在 CNN 中的首次探索；在 2014~2015 年 attention 机制在机器翻译上开始流行；2015 年 attention 机制与 RNN 结合的神经网络模型在自然语言领域广泛应用；2015—2016 年 attention 机制和 CNN 相结合的神经网络模型成为研究热点。随后，研究人员利用 attention 机制和神经网络模型进行文本情感分类层出不穷，并成为近年来的研究热点。

随着 LSTM、CNN、attention 机制等混合神经网络模型在文本情感分类中的广泛应用，在 LSTM 与 attention 机制的结合的模型在文本方面取得的卓越成绩后，人们逐渐把目光投向 attention 机制与 BiLSTM、Bi-GRU 的研究。国内外近两三年此类研究也开始涌现。田生伟等人[24]利用 Bi-LSTM 和 attention 机制结合的混合神经网络对维吾尔语事件时序关系进行识别，这种神经网络混合的方式先把文本向量放入 Bi-LSTM 层，提取一些文本信息，然后放入 attention 机制层进行深层文本特征提取，最后放入到 softmax 层进行文本情感分类；司念文等人[9]提出 LSTM-BiLSTM-Attention 网络模型更加有效地对中文词性进行标注，首先利用 LSTM 层进行文本特征提取，然后利用 BiLSTM 双方向的原理通过双向的词性来对目标词进行标注，最后利用 Attention 层进一步提取重点信息，增加词性标注的准确性；成璐[25]提出了基于 attention 机制和 BiLSTM 的神经网络模型对中文商品评论进行情感分析，将分词后的词向量的形式放入 Bi-LSTM 进行文本特征提取，放入 Attention 层突出文本分类中的重点信息；Rozental 等人[26]提出使用 BiGRU 神经网络和卷积的最大池化进行多次操作的混合模型来提取文本特征信息，最后在 softmax 层进行文本分类，用英语和西班牙语来测试实验模型，取得第一和第三的好成绩；Chen 等人[27]使用 Bi-LSTM 和位置 Attention 机制结合的混合神经进行文本分类，也取得不错的分类效果；Kumar 等人[28]使用简单的 BiLSTM 和两层 attention 机制简单混搭的模型，在 BiLSTM 层提取了文本特征信息后，分别顺序建立了词语级的 attention 机制和句子级的 attention 机制，将该实验模型在 SemEval 2017 Task 5 上进行评测，在 sub-tracks 1 和 2 的评测结果比目前最好的系统分别提高 1.7 和 3.7 个百分点。

无论是对维吾尔族语事件时序关系的分析、中文词性的分析或者商品评论分析还是国外对 BiLSTM 或 Bi-GRU 的一些最新研究，本质上都是利用神经网络和 attention 机制结合的网络模型对文本进行相关处理，提取深层次文本信息；大部分采用的是 BiLSTM 或者 BiGRU 和 attention 机制结合的混合模型，且大部分模型都经历了将文本向量放入 Bi-LSTM 层特征提取、注意力机制层突出重点信息，softmax 层文本分类三个阶段。这

些研究都把很大的注意力放在了 Bi-LSTM 或 BiGRU 神经网络和 attention 结合的模型研究上，并且取得了不错的效果。本文的算法模型在以前研究成果基础上，提出了一种更加简单的模型，采用结构更加简练、计算更加简单和存储空间更小的 BiGRU 层、attention 层、softmax 层和全连接层的简单混合神经结构来提取特征信息，进行文本分类。

2 相关技术及 BiGRU-Attention 模型

2.1 相关技术

2.1.1 长短期记忆神经

RNN 是一种用来处理序列数据的特殊的神经网络，这种神经网络的功能与人的一些思考习惯类似。想象这样一个情景，当我们做英语四、六级选词填空时，确定空格要填的单词，我们不仅要阅读空格前面的单词还要阅读前面单词的前面单词如 The sea water in the deep sea is very ___，通过阅读 very 及之前的 in the deep sea 可以知道此处应选 deep。同样的 RNN 就是为了解决文本中当前输出与当前输入和前一时刻输出的关系而诞生的神经网络。RNN 主要应用在语音识别、机器翻译、中文分词和词性标注等序列化数据的领域，并在该领域取得了良好的成果。随着研究活动的不断开展，RNN 被广泛应用在文本处理方面；近年来，被用在情感分析领域的模型更是层出不穷。经典的 RNN 结构由输入层、隐含层和输出层组成。RNN 可以学习文本上下文的语义信息，抽取到的特征信息可以作为其他神经网络或其他模型的输入，也可以放入 *softmax* 函数层直接进行情感极性的分类。RNN 在 t 时刻展开图如图 1 所示。

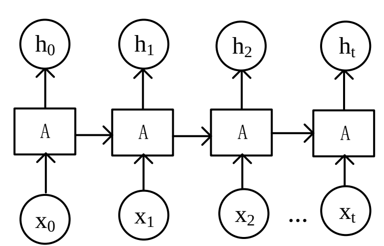


图 1 t 时刻 RNN 结构的展开图

RNN 基本的算法思想是随时间反向传播算法，但在随时间反向传播过程中，跨时间步和长时间学习使后续节点的梯度往往不能按照初值传到最初的位置，容易出现梯度弥散问题。为了克服梯度弥散的缺点,RNN 的众多变体被提出，其中 LSTM 就是 RNN 变体中一种广泛应用的经典变体。LSTM 单个神经元的具体结构如图 2 所示。

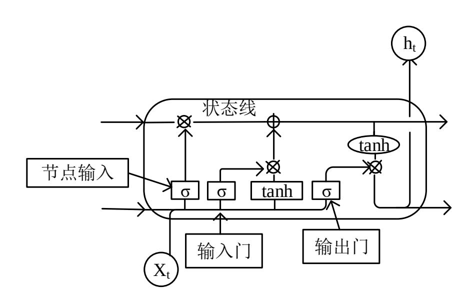


图 2 LSTM 神经元结构图

LSTM 的具体工作原理可以通过以下几个公式进行理解：

$$f_t = sigmoid\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = sigmoid\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2}$$

$$C_t = tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{4}$$

$$o_t = sigmoid\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t * tanh\left(C_t\right) \tag{6}$$

从图 2 和公式可以看出，LSTM 由输入门、记忆单元、输出门和遗忘门四部分组成。其中，x_t 表示 t 时刻的输入向量，h_{t-1} 表示上一个时刻的输出向量，$\{W_f, W_i, W_c, W_o\}$ 表示各个相应部分的权重系数矩阵，$\{b_f, b_i, b_c, b_o\}$ 表示各个相应部分的偏移向量，*sigmoid* 表示激活函数，式(1)计算的是遗忘门的值，看有多少信息可以进行保留，由式(1)的形式可以看出 t 时刻遗忘门的值由 x_t 和 h_{t-1} 共同决定；式(2)计算的是用 *sigmoid* 函数去激活的 $(W_i \bullet [h_{t-1}, x_t] + b_i)$ 细胞状态的值，式(3)计算的是由 h_{t-1} 和 x_t 决定的候选记忆单元的值;式(4)计算的是记忆状态单元通过 C_{t-1} 和 C_t 对 f_t 和 i_t 的调节作用后的值；式(5)(6)计算的是 t 时刻由 h_{t-1} 和 x_t 决定的经过内部循环和更新的 LSTM 最后的隐层状态的输出 h_t 。

2.1.2 GRU

随着 LSTM 在自然语言处理特别是文本分类任务的广泛应用，人们逐渐发现 LSTM 具有训练时间长、参数较多、内部计算复杂的缺点。Cho 等人在 2014 年进一步提出了更加简单的、将 LSTM 的单元状态和隐层状态进行合并的、还有一些其他的变动的 GRU 模型。GRU 模型是一种保持了 LSTM 效果、具有更加简单的结构、更少的参数、更好的收敛性的模型。GRU 模型由更新门和重置门两个门组成。前一个时刻的输出隐层对当前隐层的影响程度由更新门控制，更新门的值越大说明前一时刻的隐层输出对当前隐层的影响越大；前一时刻的隐层信息被忽略的程度由重置门控制，重置门的值越小说明忽略得越多。GRU 结构更加精简，具体如图 3 所示。

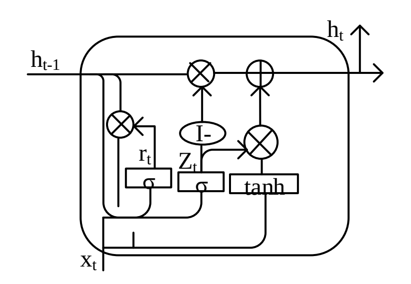


图 3 GRU 神经元结构图

GRU 模型的更新方式如下:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$h_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] \tag{7}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * h_t$$

其中：r_t 表示 t 时刻的重置门，z_t 表示 t 时刻的更新门，h_t 表示 t 时刻的候选激活状态，h_t 表示 t 时刻的激活状态，h_{t-1} 表示(t-1)

时刻的隐层状态。更新门 z 由当前状态需要被遗忘的历史信息和接受的新信息决定；重置门 r 由候选状态从历史信息中得到的信息决定。

2.1.3 BiGRU

在单向的神经网络结构中，状态总是从前往后输出的。然而，在文本情感分类中，如果当前时刻的输出能与前一时刻的状态和后一时刻的状态都产生联系如对这样一个句子进行选词填空 The sea water in the deep sea is so ______that the sun does not shine.通过在 The sea water in the deep sea 和 the sun does not shine 两部分内容，我们可以更加确信此处选填 deep，这样更有利于文本深层次特征的提取，这就需要 BiGRU 来建立这种联系。BiGRU 是由单向的、方向相反的、输出由这两个 GRU 的状态共同决定的 GRU 组成的神经网络模型。在每一时刻，输入会同时提供两个方向相反的 GRU，而输出则由这两个单向 GRU 共同决定。BiGRU 的具体结构如图 4 所示。

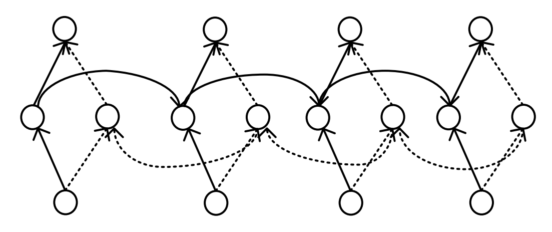


图 4　双向 GRU 结构模型图

从图 4 可以看出，BiGRU 当前的隐层状态由当前的输入 x_t、$(t\text{-}1)$时刻向前的隐层状态的输 $\overrightarrow{h_{t-1}}$，和反向的隐层状态的输出 $\overleftarrow{h_{t-1}}$ 三个部分共同决定。由于 BiGRU 可以看做两个单向的 GRU，所以 BiGRU 在 t 时刻的隐层状态通过前向隐层状态 $\overrightarrow{h_{t-1}}$ 和反向隐层状态 $\overleftarrow{h_{t-1}}$ 加权求和得到：

$$\overrightarrow{h_t} = GRU\left(x_t, \overrightarrow{h_{t-1}}\right)$$

$$\overleftarrow{h_t} = GRU\left(x_t, \overleftarrow{h_{t-1}}\right) \tag{8}$$

$$h_t = w_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t$$

其中：GRU( )函数表示对输入的词向量的非线性变换，把词向量编码成对应的 GRU 隐层状态。w_t、v_t 分别表示 t 时刻双向 GRU 所对应的前向隐层状态 $\overrightarrow{h_t}$ 和反向隐状态 $\overleftarrow{h_t}$ 所对应的权重，b_t 表示 t 时刻隐层状态所对应的偏置。

2.1.4 Attention model

Attention 机制在语音识别、机器翻译和词性标注等序列化数据中表现非凡，attention 机制可以单独使用，也可以在其他混合模型中作为其他混合模型的层使用，可以放在文本向量输入层之后也可以放在其他网络模型训练数据之后，通过对数据进行自动加权变换，把两个不同的部分联系起来，突出重点的词语，使整个系统表现出更好的性能。Attention 机制类似人脑的观察某些事物的原理，如人们为了描述某些画的内容而去观察某幅画，首先会观察这幅画上的题的字，然后根据判断有目的地去观察这幅图中表现主题的那部分内容；当去描述这幅画时，往往会先描述与这幅画最相关的内容，然后再去描述其他方面的内容；attention 机制就是一种通过在关键信息上分配足够的关注、突出局部重要信息的机制。attention 机制通常可以分为两类：时间 attention 机制和空间 attention 机制，本文用到的主要是时间注意力。attention 机制是一种类似人脑的注意力资源分配机制，通过概率权重分配的方式，计算不同时刻词向量的概率权重，使一些词语能够得到更多的关注，从而提高该隐藏层特征提取的质量。Attention model 基本结构如图 5 所示。

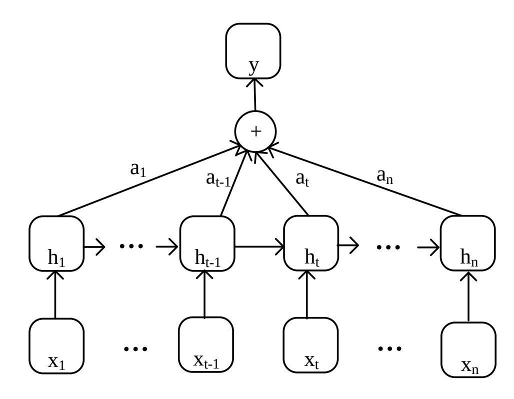


图 5　attention model 基本结构

在 attention model 中，初始隐层状态到新的隐层状态的向量 s 是各个隐层状态在新的隐层状态所占的比重大小的权重系数 α_i 与初始输入的各个隐层状态 h_i 乘积的累加和，计算公式如下：

$$s = \sum_{i=1}^{t} \alpha_i h_i$$

$$\alpha_i = \frac{exp(e_i)}{\sum_{j=1}^{n}(e_j)} \tag{9}$$

$$e_i = v_i tanh\left(w_i h_i + b_i\right)$$

其中：e_i 表示第 i 时刻隐层状态向量 h_i 所决定的能量值，v_i 和 w_i 表示第 i 时刻的权重系数矩阵，b_i 表示第 i 时刻相应的偏移量。通过式(9)可以实现从输入的初始状态到新的注意力状态的转换。

2.2　BiGRU-Attention 模型

BiGRU-Attention 模型共分为三部分：文本向量化输入层、隐含层和输出层。其中，隐含层由 BiGRU 层、attention 层和 Dense 层三层构成。BiGRU-Attention 模型结构如图 6 所示。

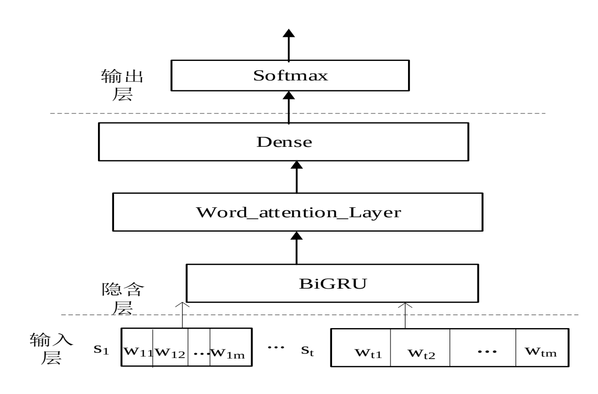


图 6　BiGRU-Attention 模型图

下面对这三层的功能分别进行介绍：

1)输入层

输入层即文本向量化输入层主要是对IMDB电影评论的25 000条数据的预处理即把这些评论数据处理成BiGRU层能够直接接收并能处理的序列向量形式。m 个单词组成 l 个句子的文本 a 即 $a=\{s_1,s_2,\cdots,s_l\}$，样本中的第 j 个句子表示为 $s_j=\{w_{j1},w_{j2},\cdots,w_{jm}\}$，进行文本向量操作，使 $w\in w^a$。文本向量化具体操作步骤如下：

a)读取数据并进行数据清洗；

b)将数据向量化为规定长度 1 000 的形式（句子长度小于规定值的，默认自动在后面填充特殊的符号；句子长度大于规定值的，默认保留前 1 000 个，多余部分截去。）

c)随机初始化数据，按 8:2 划分训练集和测试集；

d)将数据向量化后，每一条电影评论都变成了统一长度的索引向量，每一个索引对应一个词向量。

经过上面的四步的操作之后，输入的 IMDB 数据就变成根据索引对应词向量的形成词矩阵，即设处理后词向量的统一长度为 1 000，使用 glove.6B.100d 的 100 维向量的形式，在 glove.6B.100d 中不能查找到的词向量随机初始化。设 c_{ji} 为第 j 个句子的第 i 个词向量，则一条长度为 1000 的 IMDB 评论数据可以表示为：

$$c_{j1:j1000}=c_{j1}\oplus c_{j2}\oplus\cdots\oplus c_{j1000} \tag{10}$$

其中：$\oplus$表示词向量与词向量的连接操作符，$c_{j1:j1000}$ 表示为 c_{j1}，c_{j2}，$\cdots$，c_{j1000} 即为第 j 个句子的词向量矩阵。把 IMDB 每一条评论中的每一个词按照索引去对应 glove.6B.100d 中词向量，生成词向量矩阵。

2）隐含层

隐含层的计算主要分为两个步骤完成：

a)计算 BiGRU 层输出的词向量。文本词向量为 BiGRU 层的输入向量。BiGRU 层的目的主要是对输入的文本向量进行文本深层次特征的提取。根据 BiGRU 神经网络模型图，可以把 BiGRU 模型看做由向前 GRU 和反向 GRU 两部分组成，在这里简化为式(11)。在第 i 时刻输入的第 j 个句子的第 t 个单词的词向量为 c_{ijt},通过 BiGRU 层特征提取后，可以更加充分地学习上下文之间的关系，进行语义编码，具体计算公式如式(11)所示。

$$h_{ijt}=BiGRU\left(c_{ijt}\right),\ t\in[1,\ m] \tag{11}$$

b)计算每个词向量应分配的概率权重。这个步骤主要是为不同的词向量分配相应的概率权重，进一步提取文本特征，突出文本的关键信息。在文本中，不同的词对文本情感分类起着不同的作用。地点状语、时间状语对文本情感分类来说，重要程度极小；而具有情感色彩的形容词对文本情感分类却至关重要。为了突出不同词对整个文本情感分类的重要度，BiGRU-Attention 模型中引入了 attention 机制层。Attention 机制层的输入为上一层中经过 BiGRU 神经网络层激活处理的输出向量 h_{ijt}，attention 机制层的权重系数具体通过以下几个公式进行计算：

$$u_{ijt}=\tanh\left(w_w h_{ijt}+b_w\right)$$
$$\alpha_{ijt}=\frac{exp\left(u_{ijt}^T u_w\right)}{\sum_t exp\left(u_{ijt}^T u_w\right)} \tag{12}$$
$$s_{ijt}=\sum_{i=1}^{n}\alpha_{ijt}h_{ijt}$$

其中: h_{ijt} 为上一层 BiGRU 神经网络层的输出向量，w_w 表示权重系数，b_w 表示偏置系数，u_w 表示随机初始化的注意力矩阵。Attention 机制矩阵由 attention 机制分配的不同概率权重与各个隐层状态的乘积的累加和，使用 softmax 函数做归一化操作得到。

3)输出层

输出层的输入为上一层 attention 机制层的输出。利用 softmax 函数对输出层的输入进行相应计算的方式从而进行文本分类，具体公式如下：

$$y_j=softmax\left(w_1 s_{ijt}+b_1\right) \tag{13}$$

其中: w_1 表示 attention 机制层到输出层的待训练的权重系数矩阵，b_1 表示待训练相对应的偏置，y_j 为输出的预测标签。

2.3 BiGRU-Attention 模型训练方法

本文 BiGRU-Attention 模型以 IMDB 数据集、预设参数和迭代次数 N 为输入，经过文本向量化输入层把 IMDB 数据集处理词向量的形式，利用 BiGRU-Attention 模型对 IMDB 数据集进行分类。

算法　BiGRU-Attention 神经网络的文本情感分类算法

输入: IMDB 数据集、预设参数、迭代次数 N。

输出: IMDB 数据集的情感分类。

在文本向量化输入层对数据进行清洗、划分为固定长度、随机初始化、划分训练集和测试集、转化成对应词向量。

IMDB 数据集对应的词向量的形式，设文本中某个词对应的词向量为 x_t。

对 IMDB 每一条电影评论的处理：

for hop=1 to h:

$$\overrightarrow{h_t}=GRU\left(x_t,\overrightarrow{h_{t-1}}\right)$$

$$\overleftarrow{h_t}=GRU\left(x_t,\overleftarrow{h_{t-1}}\right)$$

$$h_t=w_t\overrightarrow{h_t}+v_t\overleftarrow{h_t}+b_t$$
$$u_{ijt}=\tanh\left(w_w h_{ijt}+b_w\right)$$
$$\alpha_{ijt}=\frac{exp\left(u_{ijt}^T u_w\right)}{\sum_t exp\left(u_{ijt}^T u_w\right)}$$

$$s_{ijt}=\sum_{i=1}^{n}\alpha_{ijt}h_{ijt}$$

end for

使用 softmax 函数计算分类的结果概率 $y_j=softmax\left(w_1 s_{ijt}+b_1\right)$ 并与原始的标签$\hat{y}$进行对比，本文的目标函数为

$$\text{loss}=-\sum_j \hat{y}^{(j)}logy^{(j)} \tag{14}$$

从式(14)可知，通过上述的训练步骤，对从 1 到 h 个单词分别进行特征提取，分配相应权重累加求和,Dense 层进一步提取特征，最后在 softmax 输出层进行分类，然后对所分类的各个影评标签值和 $logy^{(i)}$ 相乘的累加和，累加和为负值，取相反，使损失最小，计算误差，使用 rmsprop 作为训练器，使模型训练和收敛的更快，在沿随时间反向误差传播过程中根据误差不断调整并更新权重和偏置，直到达到迭代次数或小于某一个固定的精度为止。

3　实验

3.1　实验设置

为了验证 BiGRU-Attention 模型的有效性，实验选用公开的 IMDB 作为数据集。训练集和测试集按 8:2 进行划分，分别用于模型训练和模型测试，具体划分情况如表 1 所示。

表 1　数据集

数据集	数据集	语句/条	类别
IMDB	训练集	20000	2
	测试集	5000	2

本实验采用准确率、损失率和迭代时间作为实验的评价标准。数据预处理是实验数据放入神经网络训练前最重要的一步。文本向量层已经对数据处理做了详细的论述，在这里不做陈述。本实验采用 keras 深度学习框架，底层为 TensorFlow，使用 Python 语言编程实现；实验运行环境为 JetBrains PyCharm 软件、Win10 系统、内存 8 GB 等。本实验模型为文本向量输入层、隐藏层和文本分类层的三层网络结构。实验中有许多超参数需要设置和调整，超参数设置和调整都是在每一次迭代完成后，根据实验的准确率、损失率调整的。经过多次迭代，实验设置的超参数如表 2。

表 2　模型参数设置

参数	值
BiGRU 隐藏层节点数	100
Loss 函数	Categorical_crossentropy
optimzer	rmsprop
Batch_size	50
词向量维度	100

3.2　实验评价标准

本文采用准确率、损失率和迭代时间作为实验的评价标准，设样本数总数为 M，被正确分类的样本数为 m,则正确率(Accuracy)为

$$\text{Accuracy} = \frac{m}{M} \tag{15}$$

损失函数是在每一次随机批量训练的过程中，按照式(14)的批量样本的相乘累加和取反得到。

迭代时间是 10 次迭代过程每一次时间相加的和取均值得到的。

3.3　实验步骤

BiGRU-Attention 模型的具体实验步骤如下：

a)进行数据清洗；

b)统一为固定长度的索引向量的形式；

c)划分训练集和测试集；

d)每一个索引向量对应词向量；

e)拼接词向量，组成词矩阵，形成 BiGRU-Attention 模型的输入；

f)把文本输入矩阵放入 BiGRU 层，采用公式(11)进行模型训练，提取文本层次特征；

g)放入 Attention 模型给文本的向量分配相应的权重；

h)采用 IMDB 测试集进行 BiGRU-Attention 模型性能评估。

3.4　对比实验

BiGRU-Attention 模型与以下常见的 7 模型进行对比。7 模型的结构情况如下：

a)ASC。Rozental 等人[26]提出的利用 Bi-LSTM 进行特征提取和降维、最大池化层突出重点信息、全连接层降维，采用四个不同维度处理的向量进行连接，进入全连接层，最后 *softmax* 分类的网络模型

b)DBLSTM-Attention。Chen 等人[27]提出了 BiLSTM 和位置 Attention 机制结合的混合实验模型。

c)BiLSTM-Attentions。Kumar 等人[28]提出的利用 BiLSTM 和两层 Attention 机制结合的混合实验模型。

d)HDBN 模型。Yan 等人[29]提出了主要利用 DBM 降噪降维、DBN 提取层次特征和 *softmax* 层文本分类的将 DBM 和 DBN 混合的 HDBN 的混合神经网络模型。

e)S-LSTM。Zhu 等人[30]提出的利用树形结构的 LSTM 模型并通过记忆细胞来记住历史信息的混合模型。

f)BiLSTM-Attention 模型。BiLSTM-Attention 模型与 BiGRU-Attention 模型属于同种类型的混合实验模型。BiLSTM-Attention 模型就是把 BiGRU-A-ttention 模型中的 BiGRU 层换成 BiLSTM 层，其他实验设置均相同。

g)LSTM-Attention 模型。LSTM-Attention 模型、Bi-LSTM-Attention 模型与 BiGRU-Attention 模型都属于同种类型的混合实验模型。LSTM-Attention 模型就是把 BiGRU-Attention 模型中的 BiGRU 层换成 LSTM 层，其他实验设置均相同。

以上所有的实验模型都基本分为文本向量输入层、隐藏层即文本信息提取功能层和文本分类层的三层网络结构，超参数相同，迭代次数均为 10 次。从以上分析可以看出，除了实验变量层即模型功能层外，四种模型其他条件均相同，保证了实验变量层的唯一性和实验结果的针对的比较性。

3.5　实验结果分析

3.5.1 与 HDBN、S-LSTM 、ASC 、DBLSTM-Attention 、BiLSTM-Attentions 模型的对比实验

本实验选用在 10 次迭代过程中，在测试集上最高的准确率

作为该模型的准确率，对应的测试集最高准确率的损失率即为该模型的损失率，对应的迭代时间即为该模型的迭代时间记为 time，具体如表 3 所示。

表 3　不同模型的分类结果

模型	准确率	损失率	time
BiGRU-Attention	90.54%	0.2430	1100s
ASC	89.96%	0.2445	985s
BiLSTM-Attentions	90.34%	0.2438	1417s
DBLSTM-Attention	88.85%	0..2650	1345s
S-LSTM	88.74%	0.2847	1205s
HDBN	86.94%	0.3947	1115s

从表 3 可以看出，本文所提出的 BiGRU-Attention 模型在准确率、损失率方面的表现性能均优于其他五种实验模型，时间代价高于 ASC 实验模型但低于其他四种模型。从总体上看，这六种实验模型的准确率、损失率、时间代价比较接近，准确率、损失率、时间代价总是随着模型复杂度有所变化、上下波动。表格从上往下进行依次对比，可以看出 BiGRU-Attention 模型准确率和损失率性能优于 ASC 模型，两模型都利用了 BiGRU 神经网络层，来提取文本的重点信息；两模型的区别在于一个是在 BiGRU 层的基础上加了 Attention 层，利用 Attention 层分配相应的权重来突出文本的重点特征，而另一个则是在 BiGRU 层的基础上增加了最大池化层，利用最大池化的突出重要信息，遗忘次要信息的特征。由两者的比较可以看出，Attention 层在突出重要信息方面的表现性能要优于卷积的最大池化层，表现在实验评价指标性能方面就是提高准确率，降低损失率，但因为 Attention 层突出信息的功能是通过不断加权计算进行的，相比于单纯的最大池化，时间代价增加；BiGRU-Attention 模型在实验的评价指标方面优于 DBLSTM-Attention，这两个实验模型均存在 attention 层，attention 层的主要功能基本相同，模型结构也基本相同；不同在于实验模型的第一层一个是 BiGRU 层，而另一个是 BiLSTM 层，通过本实验这两个实验模型的评价指标观察分析，可以看出 BiGRU 神经网络在提取文本重点信息、减少计算量方面相比 BiLSTM 神经网络表现更优。BiLSTM-Attentions 模型与 DBLSTM-Attention 模型进行比较，两者的区别在于 BiLSTM-Attentions 模型要比 DBLSTM-Attention 模型多一个 Attention 层，通过对实验评价指标的分析，我们可以清楚的看出 Attention 层突出文本信息的重要作用。

总之，通过对上面几个模型的对比分析可以得出这样的一个结论：在 IMDB 数据集上，BiGRU 模型要优于 BiLSTM，因为 BiGRU 收敛速度更快，计算更简单，参数更少，在提高准确率降低损失率的同时降低模型的训练时间；attention 层在突出重点信息方面要优于卷积的池化层。所以 BiGRU-Attention 模型在实验的指标方面均表现比较优异。

3.5.2 迭代次数

本实验选用 10 次迭代实验来作为实验的评价指标分析。在 10 次迭代实验中，准确率、损失率和时间代价都不是一成不变的，都在动态变化着。为了深刻地反映这种动态变化，本文在所有的实验模型中，选取了最具典型、与本实验模型最相近、且最新提出的四种（BiGRU-Attention、ASC、BiLSTM-Attentions 和 DBLSTM-Attention）模型进行对比。

a) 本文将 BiGRU-Attention 模型与 ASC、BiLSTM-Attentions、DBLSTM-Attention 模型在训练集上进行训练，在测试集上进行对比实验，得到在测试集上的准确率与迭代次数之间的关系，具体如图 7 所示。

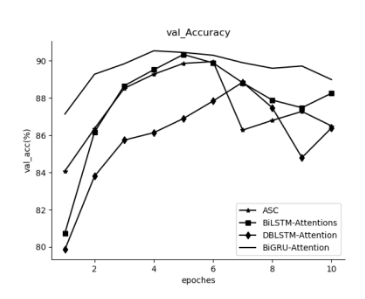


图 7　四种模型准确率变化情况

图 7 展示了四种模型随迭代次数准确率的变化情况，各个模型从下往上看，总体上准确率不断地提高，BiGRU-Attention 模型的准确率始终高于其他三种模型，其中 BiGRU-Attention 模型的准确率高于 ASC 模型，突出 Attention 层相比最大池化更具快速凸显的重要信息、提取的文本深层次特征，快速收敛，快速提高准确性率的特征，在第三次迭代时就能达到最高准确率，并且在最初训练时的准确率就高于其他三种，训练效果较好；BiGRU-Attention 模型总体上准确率平稳变化，但在某些迭代次数上可能会出现准确率较低的情况，但准确率始终高于其他三种模型；从总体上看，BiGRU-Attention、ASC 和 BiLSTM-Attentions 这三种模型曲线比较接近，而 DBLSTM 曲线相对变化起伏比较大。可见，BiGRU 模型在提取文本深层次特征方面表现更优且比较稳定；从迭代次数上看，并不是迭代次数越多，准确率越高，每一种模型都有其达到最高准确率的最佳的迭代次数如 BiGRU-Attention 模型在第四次迭代时准确率最高，而 ASC 模型在第六次迭代时准确率最高。从以上分析可以看出，BiGRU-Attention 模型能够在迭代次数最少的情况下有效地提高训练数据的准确率。

b)实验完成一次迭代所需要的时间是实验所需的时间代价。图 8 给出了四种模型在相同的实验条件下完成一次迭代所需要时间的变化趋势曲线。

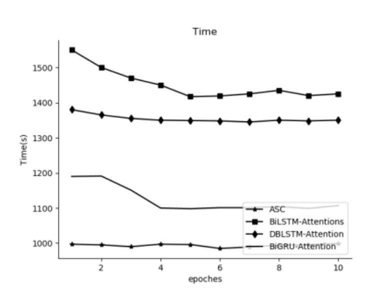


图 8　四种模型的迭代时间图

如图 8 可以看出，各个模型的迭代时间总体上没有多大起伏，整体时间趋于平稳，一般都经过了最低的迭代时间之后，再次训练的时候，训练时间不会再有较大的起伏；ASC 模型完成一次迭代训练的时间最短，这与卷积的最大池化层的训练迭代收敛速度是分不开的；BiGRU-Attention 模型的迭代时间曲线位于 DBLSTM-Attention 迭代时间曲线和 ASC 曲线之间，总体时间性能居中，BiGRU-Attention 和 DBLSTM-Attention 这两个模型的区别基本上在于 BiLSTM 和 BiGRU 的不同所引起的，二者迭代时间的不同，说明：迭代速度不同，BiGRU 比 BiLSTM 模型具有计算更快、参数更少的特性；BiLSTM-Attentions 模型的迭代时间曲线相对来说最高，是因为 BiLSTM 神经网络相对计算较复杂、增加了计算的时间并且 Attention 层在突出重点信息的同时也增加了加权计算时间的缘故。

c)测试数据集的损失率也是衡量模型性能的一个重要的标准，损失率越小说明模型的性能指标越好。从图 9 可以看出，各个模型在各次迭代过程中，有些模型损失率的起伏还是比较大的如 BiGRU-Attention 、DBLSTM-Attention 和 BiLSTM-Attentions 模型损失率曲线起伏比较大，但 ASC 模型达到最低的损失率后，模型曲线没有多大变化，说明没有加入 Attention 层的模型的稳定性比较好；各个模型达到损失率最低的迭代次数也不尽相同，比如 BiGRU-Attention 模型在第四次迭代时达到最低损失率 0.2430；而 ASC 模型在第六次迭代时达到最低的损失率 0.2445；BiGRU-Attention 模型最初迭代的时候损失率明显的低于其他模型，达到最小的损失率后随着迭代次数的增加，其损失率呈现逐渐上升的趋势;而 ASC 模型总体曲线变化相对比较平稳，总体呈下降趋势。

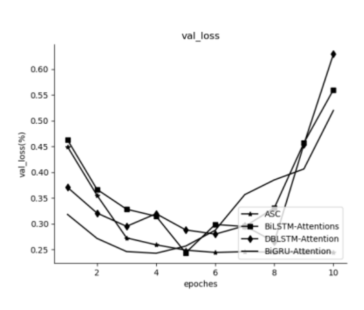


图 9　训练集损失率变化曲线

结合图 7~9 可以看出，虽然 BiGRU-Attention 模型相比 ASC 模型需要训练的时间代价要稍微地高一些，但其准确率更高，损失率更低，达到最佳准确率和损失率所需要的迭代次数更少，总体上需要的训练时间较少；BiGRU-Attention 模型在训练集上无论准确率、损失率还是在迭代次数上都具有相比 DBLSTM-Attention 和 BiLSTM-Attentions 模型无法比拟的优势。综上，本实验证明了 BiGRU-Attention 的模型有效性。

3.5.3 与 BiLSTM-Attention 模型、LSTM-Attention 模型的对比实验

虽然 BiGRU-Attention 模型与 HDBN、S-LSTM 、ASC 、DBLSTM-Attention 、BiLSTM-Attentions 模型的对比实验已经能充分的说明，BiGRU-Attention 模型具有降低模型训练时间、高准确率和低损失率的特点。但是毕竟不属于同种类型混合实验模型的对比，缺少实验的说服力。因此，选择 BiGRU-Attention 模型与 BiLSTM-Attention 模型、LSTM-Attention 模型的对比实验，具体情况如表 4 所示。

表 4　三种模型对比

模型	准确率	损失率	time
BiGRU-Attention	90.54%	0.2430	1100s
BiLSTM-Attention	89.23%	0.2460	1352s
LSTM-Attention	87.63%	0.2520	1221s

通过三种模型的准确率、损失率和迭代时间的对比，可以看出 BiGRU-Attention 模型在训练集和测试集比其他两种实验模型表现更优。

4　结束语

本文提出了一种新的基于 BiGRU-Attention 的神经网络模型，比起目前最广泛使用的 BiLSTM 神经网络和 Attention 的混合模型，能够在提高准确率的同时降低损失率，适当地降低了模型的训练时间。一方面说明 BiGRU 比 BiLSTM 更加的简单，训练时间要快一些；另一方面说明 BiGRU 与 Attention 模型结合的有效性；在与最新提出来的模型相比，总体上看来，BiGRU-Attention 的神经网络模型略胜其他网络模型。虽然基于 BiGRU-Attention 的神经网络模型在 IMDB 数据集上的准确率较高，损失率较低，但随着数据量的增大，其准确率会有所下降；加入 Attention 机制模型需要进一步自动加和处理所有的对象并且要存储相应的权重信息，增加了系统的计算和系统开销。下一步将寻找能够在数据量巨大时，训练时间较短，准确率较高，损失率较低，计算量小，系统开销小，更适合文本情感分类的神经网络模型。

参考文献：

[1] Lecun Yann, Bottou Lenon, Bengio Yoshua, *et al.* Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[2] 陈珂，梁斌，柯文德，等. 基于多通道卷积神经网络的中文微博情感分析 [J]. 计算机研究与发展, 2018, 55 (5): 945-957. (Chen Ke, Liang Bin, Ke Wende, *et al.* Chinese micro-blog emotional analysis based on multi-channel convolutional neural network [J]. Computer Research and Development, 2018, 55 (05): 945-957.)

[3] Mikolov T, Sutskever I, Chen Kai, *et al.* Distributed representations of words and phrases and their compositionality [J]. Neural Information Processing Systems, . 2013: 3111-3119.

[4] 梁军，柴玉梅，原慧斌，等. 基于深度学习的微博情感分析 [J]. 中文信息学报, 2014, 28 (5): 155-161. (Liang Jun, Chai Yumei, Yuan Huibin, *et al.* Micro-blog emotional analysis based on deep learning [J]. Chinese Journal of information, 2014, 28 (5): 155-161.)

[5] 白静，李霏，姬东鸿. 基于注意力的 BiLSTM-CNN 中文微博立场检测

模型 [J]. 计算机应用与软件, 2018, 35 (3): 266-274. (Bai Jing, Li Fei, Ji Donghong. Attention based BiLSTM-CNN Chinese micro-blog stance detection model [J]. Computer Application and Software, 2018, 35 (3): 266-274.)

[6] 梁斌，刘全，徐进，等. 基于多注意力卷积神经网络的特定目标情感分析 [J]. 计算机研究与发展，2017，54 (8): 1724-1735. (Liang Bin, Liu Quan, Xu Jin, *et al.* Specific target emotion analysis based on multi-attentions convolution neural network [J]. Computer Research and Development, 2017, 54 (8): 1724-1735.)

[7] 赵勤鲁，蔡晓东，李波，等. 基于 LSTM-Attention 神经网络的文本特征提取方法 [J]. 现代电子技术，2018，41 (8): 167-170. (Zhao Qinlu, Cai Xiaodong, Li Bo, *et al.* Text feature extraction method based on LSTM-Attention neural network [J]. Modern Electronic Technology, 2018, 41 (8): 167-170.)

[8] 杨东，王移芝. 基于 Attention-based C-GRU 神经网络的文本分类 [J]. 计算机与现代化，2018 (2): 96-100. (Yang Dong, Wang Yizhi. Text classification based on Attention-based C-GRU neural network [J]. Computer and Modernization, 2018 (2): 96-100.)

[9] 司念文，王衡军，李伟，等. 基于注意力长短时记忆网络的中文词性标注模型 [J]. 计算机科学，2018，45 (4): 66-70+82. (Si Nianwen, Wang Hengjun, Li Wei, *et al.* Chinese POS tagging model based on attention time memory network [J]. Computer Science, 2018, 45 (4): 66-70+82.)

[10] 王业沛，宋梦姣，王譞，等. 基于深度学习的判决结果倾向性分析 [J/OL]. 计算机应用研究，2019，36 (2) . [2018-06-23]. http://kns. cnki. net/kcms/detail/51. 1196. TP. 20180208. 1753. 154. html. (Wang Yepei, Song Meijiao, Wang Pu, *et al.* Analysis of decision results propensity based on deep learning [J/OL]. Computer Application Research, 2019, 36 (2) . [2018-06-23]. http://kns. cnki. net/kcms/detail/51. 1196. TP. 20180208. 1753. 154. html.)

[11] 朱星嘉，李红莲，吕学强，等. 一种改进的 attention-based LSTM 特征选择模型 [J]. 北京信息科技大学学报：自然科学版，2018，33 (2): 54-59. (Zhu Xingjia, Li Honglian, Lu Xueqiang, *et al.* An improved Attention-Based LSTM feature selection model [J]. Journal of Beijing University of Information Technology: Natural Science Edition, 2018, 33 (2): 54-59.)

[12] 李阳辉，谢明，易阳. 基于深度学习的社交网络平台细粒度情感分析 [J]. 计算机应用研究，2017，34 (3): 743-747. (Li Yanghui, Xie Ming, Yi Yang. Fine-grained Emotional Analysis of Social Networking Platform Based on Deep Learning [J]. Computer Applications, 2017, 34 (3): 743-747.)

[13] 刘洋. 基于 GRU 神经网络的时间序列预测研究 [D]. 成都：成都理工大学，2017. (Liu Yang. Time series prediction based on GRU neural network [D]. Chengdu: Chengdu University of Technology, 2017.)

[14] 李骁，黄征. 基于 GRU 网络的互联网信息挖掘 [J]. 信息技术，2018 (3): 1-5, 9. (Li Xiao, Shawn Huang. Internet Information Mining Based on GRU network [J]. Information Technology, 2018 (3): 1-5, 9)

[15] 黄兆玮，常亮，宾辰忠，等. 基于 GRU 和注意力机制的远程监督关系抽取 [J/OL]. 计算机应用研究，2019，36 (10) . [2018-07-05]. http://kns. cnki. net/kcms/detail/51. 1196. TP. 20180619. 1516. 012. html. (Huang Zhaowei, Chang Liang, Bin Chenzhong, *et al.* Remote supervisory relationship extraction based on GRU and Attention mechanism [J/OL]. Computer Application Research, 2019, 36 (10) . [2018-07-05]. http://kns. cnki. net/kcms/detail/51. 1196. TP. 20180619. 1516. 012. html.)

[16] 张玉环，钱江. 基于两种 LSTM 结构的文本情感分析 [J]. 软件，2018，39 (1): 116-120. (Zhang Yuhuan, Qian jiang. Text sentiment analysis based on two kinds of LSTM structure [J]. Software, 2018, 39 (1): 116-120.)

[17] 田竹. 基于深度特征提取的文本情感极性分类研究 [D]. 济南：山东大学，2017. (Tian Zhu. Text sentiment polarity classifications based on depth feature extraction [D]. Jinan: Shandong University, 2017.)

[18] Zhou Guobing, Wu Jianxin, Zhang Chenlin, *et al.* Minimal gated unit for recurrent neural networks [J]. International Journal of Automation and Computing, 2016, 13 (3): 226-234.

[19] 黄磊，杜昌顺. 基于递归神经网络的文本分类研究 [J]. 北京化工大学学报：自然科学版，2017，44 (1): 98-104. (Huang Lei, Du Changshun. Text classification based on recurrent neural network [J]. Journal of Beijing University of Chemical Technology: Natural Science Edition, 2017, 44 (1): 98-104.)

[20] Mnih Volodymyr, Heess Nicolas, Graves Alex, *et al.* Recurrent models of visual attention [J]. Neural Information Processing Systems, 2014: 2204-2212.

[21] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [C]// Proc of International Conference on Learning Representations. 2015: 1-15.

[22] Luong T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [J]. Empirical Methods in Natural Language Processing, 2015: 1412-1421

[23] Yin Wenpeng, Schutze H, Xiang Bing, *et al.* Attention-based convolutional netural network for modeling sentence pairs [J]. Transactions of the Association for Computional Linguistics, 2016, 4: 259-272

[24] 田生伟，胡伟，禹龙，等. 结合注意力机制的 Bi-LSTM 维吾尔语事件时序关系识别 [J]. 东南大学学报：自然科学版，2018，48 (3): 393-399. (Tian Shengwei, Hu Wei, Yu Long, *et al.* Bi-LSTM Uygur event sequence recognition combined with attention mechanism [J]. Journal of Southeast University: Natural Science Edition, 2018, 48 (3): 393-399.)

[25] 成璐. 基于注意力机制的双向 LSTM 模型在中文商品评论情感分类中的研究 [J]. 软件工程，2017，20 (11): 4-6, 3. (Cheng lu. A bidirectional LSTM model based on attention mechanism for emotional classification of Chinese comments [J]. Software Engineering, 2017, 20 (11): 4-6, 3)

[26] Rozental Alon, Fleischer Daniel. Amobee at SemEval-2018 Task 1: GRU neural network with a CNN attention mechanism for sentiment classification [C]// North American Chapter of the Association for Computational Linguistics. 2018: 218-225.

[27] Chen Peng, Sun Zhongqian, Bing Lidong, *et al*. Recurrent Attention Network on Memory for Aspect Sentient Analysis [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2017: 452-461.

[28] Kumar A, Kawahara D, Kurohashi S, *et al*. Knowledge-enriched two-layered attention for sentiment analysis [C]// North American Chapter of the Association for Computational Linguistics. 2018: 253-258.

[29] Yan Yan, Yin Xucheng, Li Sujiang, *et al*. Hybrid deep belief network [J]. Computational Intelligence and Neuroscience, 2015 (5): 650527: 1-650527: 9.

[30] Zhu Xiaodan, Sobhani P, Guo Hongyu. Learning document semantic representation with long short-term memory over recursive structures [C]// Proc of the 32nd International Conference on International Conference on Machince Learning. 2015: 1604-1612.